Water Resources Research

RESEARCH ARTICLE
10.1029/2020WR028179

Nonlinearity and Multivariate Dependencies in the Terrestrial Leg of Land-Atmosphere Coupling

Hsin Hsu[1] and **Paul A. Dirmeyer**[1,2]

[1]George Mason University, Fairfax, VA, USA, [2]Center for Ocean-Land-Atmosphere Studies, George Mason University, Fairfax, VA, USA

Key Points:

- Nonlinear and multivariate dependencies in land-atmospheric coupling are diagnosed using multivariate mutual information (MMI)
- Integrated MMI method decomposes drivers of surface flux variability into linear versus nonlinear; unique, redundant, and synergistic components
- Nonlinear and synergistic dependencies of surface heat fluxes on soil moisture and net radiation are widespread across the globe

Supporting Information:

- Supporting Information S1

Correspondence to:
H. Hsu,
hhsu@gmu.edu

Citation:
Hsu, H., & Dirmeyer, P. A. (2021). Nonlinearity and multivariate dependencies in the terrestrial leg of land-atmosphere coupling. *Water Resources Research*, *57*, e2020WR028179. https://doi.org/10.1029/2020WR028179

Received 20 JUN 2020
Accepted 27 JAN 2021

Abstract Most studies of land-atmosphere coupling have focused on bivariate linear statistics like correlation. However, more complex dependencies exist, including nonlinear relationships between components of land-atmosphere coupling and the transmutability of relationships between soil moisture and surface heat fluxes under different environmental conditions. In this study, a technique called multivariate mutual information, based on information theory, is proposed to quantify how surface heat fluxes depend on both surface energy and wetness conditions, that is, net radiation and soil moisture, seasonally across the globe using reanalysis data. Such interdependency is then decomposed into linear and nonlinear contributions, which are further decomposed as different components explainable as the unique contribution from individual surface conditions, redundant contributions shared by both surface conditions, and the synergistic contribution from the concurrent action of net radiation and soil moisture. In reanalysis data, the dependency linearly contributed from soil moisture bears a similar global pattern to previously identified hot spots of coupling. The linear unique contributions of net radiation and soil moisture are mainly nonoverlapping, which suggests two separate regimes are governed by either energy or water limitations. These patterns persist when the nonlinearity is superimposed, thus reinforcing the validity of the land-atmospheric coupling hot spot paradigm and the spatial division of energy-limited as well as water-limited regions. Nevertheless, strong nonlinear relationships are detected, particularly over subtropical regions. Synergistic components are found across the globe, implying widespread multidimensional physical relationships among net radiation, soil moisture, and surface heat fluxes that previously had only been inferred locally.

Plain Language Summary Most of our knowledge of how land surface states affect weather and climate around the world is based on common statistical methods that assume straight lines can be fitted to determine how evaporation and heating of the air can be controlled by soil moisture. Furthermore, studies have historically looked one-at-a-time at a single cause affecting a single response. But local studies have shown that coupling between land and atmosphere can be more complex, involving multiple factors working simultaneously and often nonlinearly. This study uses a novel combination of techniques to discern how variations in processes like evaporation are driven by multiple factors such as soil moisture and available soil and thermal energy at the land surface that can make separately unique contributions, redundant contributions, and compound contributions due to factor interaction. Each of these contributions is also broken down into simple linear and other nonlinear components, helping show how much land-atmosphere interaction the previous methods were missing, and suggesting which regions, seasons and potential physical processes need further study.

1. Introduction

Land-atmosphere coupling acting through water and energy cycles is an important component of the Earth system (Seneviratne et al., 2010). In addition to the process of precipitation directly moistening the land surface, the return process of the moistened land influencing the atmosphere is rather intricate (Eltahir, 1998). Conventionally, the mechanism behind such upward pathway, how soil moisture affects precipitation, can be divided into a terrestrial leg and an atmospheric leg (Santanello et al., 2018). The terrestrial leg of coupling describes the ability of soil moisture to affect the partitioning of surface heat fluxes. The atmospheric leg of coupling characterizes how surface heat fluxes can modify the properties of the lower atmosphere and ultimately cloud formation and precipitation. Generally, with sufficient available energy, anomalously wet soil can moisten the atmosphere by enhancing evapotranspiration, cooling the surface while also

cooling the overlying atmosphere by suppressing the release of sensible heat flux. The resulting moistened atmosphere and suppressed boundary layer growth compete to determine the net impact on cloud formation and thus precipitation.

Traditionally, studies attempting to identify regions of strong land-atmosphere coupling have used statistical frameworks with linear dependencies between two factors, of which the influences of temporal variability are hypothesized to be significant in the explored "leg" of coupling. For instance, the complete coupling that consists of a terrestrial leg and an atmospheric leg has been diagnosed by the proportion of precipitation variance explained by soil moisture using output from multiple climate models (Koster et al., 2004) during boreal summer and has shown that regions with strong land-atmosphere coupling are mostly located in the semiarid regions. A similar pattern has been detected in reanalysis data with a sensitivity index involving variances and correlations of soil moisture and surface heat fluxes (Dirmeyer, 2011). These studies have described a canonical global pattern of land-atmosphere coupling strength and thus many observational studies have explored the land surface processes over hot spots such as the North American Great Plains (e.g., Santanello et al. 2013; Tao et al., 2019), the Sahel (e.g., Los et al., 2006; Yu et al., 2017), and Australia (e.g., Herold et al., 2016; Kala et al., 2015). Since the soil moisture-surface heat flux-precipitation interactions have long been demonstrated as a potential key to improve the skill of subseasonal to seasonal forecasts, the importance of land-atmosphere interactions in such regions with strong land-atmosphere coupling is matched only by the eastern equatorial Pacific, a hot spot of ocean-atmosphere coupling associated with the phenomenon known as El Niño-Southern Oscillation (ENSO).

However, dependencies among the environmental factors within land-atmosphere coupling may be more complex than can be depicted by linear frameworks. A classic example is the long-recognized threshold behavior in which the flux of latent heat behaves dramatically differently once soil moisture crosses a certain critical value (Budyko 1963, 1974). Another source is multivariate dependence, which is revealed in a recent in situ observation-based analyses showing the relationship between soil moisture and surface heat flux is nonunique (Haghighi et al., 2018). The embedded complex dependencies that cannot be recognized through the canonical linear coupling framework motivate this study, re-examining land-atmosphere coupling strength while addressing its nonlinear and multivariate aspects.

Realizing that there are facets of nature that are unidentifiable by common linear statistical methods, studies have applied techniques based on information theory to explore nonlinear dynamics and multivariate interdependency in the ecosystem (Goodwell & Kumar, 2017a, 2017b; Neuper & Ehret, 2019; Qiu et al., 2020; Ruddell & Kumar, 2009). These localized studies as analytic paradigms have opened the possibility to apply information-theoretic statistical methods at a larger spatial scale. Here we employ an information theory-based approach to assess the terrestrial leg of land-atmosphere coupling at the global scale by quantifying both linear and nonlinear dependencies on surface heat fluxes by energy and moisture availability at the land surface, for which net radiation and soil moisture are used as proxies respectively. A brief introduction of information theory measurement is presented in Sections 2.1−2.3. Since long-term global observations of some of the required variables are not available, fields from reanalysis data are used to develop and test the analysis in this study. This is discussed in more detail in Section 2.4.

Following the commonly used terminology in information theory to describe causes and effects, soil moisture and net radiation are referred to as source variables in this study and surface heat flux is called the target variable. Dependency between the target and one source is called mutual information (MI) and dependency between the target and multiple sources is called multivariate MI (MMI). The global pattern of MMI and its seasonal cycle are calculated. Furthermore, we integrate methods of calculating nonlinearity in MI with the method of decomposing the MMI into different contributed components. This enables us also to separate MMI into linear and nonlinear components, interpretable as the unique information contributed by a particular source, the redundant information provided identically by both sources, and the synergistic information created by the interaction of the sources. The diagnosed nonlinearity and synergistic components reveal unexplored aspects of land-atmosphere coupling that can inform process understanding with potential to improve model parameterizations and prediction skill.

2. Methodology and Data

In this section, we first introduce the existing information measurements and their partitioning. We propose a combination of those measurements to decompose the information in land surface states and fluxes into their basic components. Then, the integrated application of the technique is illustrated. Lastly, the data and specifics of significance testing are described.

2.1. Information Measurement

2.1.1. Shannon Entropy

Shannon Entropy H (Shannon, 1948) quantifies the amount of uncertainty of a single random variable X with the probability distribution function $p(x)$:

$$H(X) = -\sum p(x)\log_2 p(x) \tag{1}$$

Various bases of the logarithm have been used in different applications; here base-2 is used, so the quantity is in units of bits. In this study, only the probability distribution in time is examined. For summation, $p(x)$ must be expressed across a finite number of bins—the procedure for bin selection is discussed in the Supporting Information S1.

2.1.2. Mutual Information

Conditional entropy $H(X_{\text{tar}} \mid X_s)$ is an expression of Shannon entropy that quantifies the amount of uncertainty of a target variable X_{tar} given knowledge of a single source variable X_s:

$$H(X_{\text{tar}} \mid X_s) = -\sum p(x_s, x_{\text{tar}})\log_2 \frac{p(x_s, x_{\text{tar}})}{p(x_s)} \tag{2}$$

MI (Cover & Thomas, 2006) measures the shared information between two random variables. In other words, MI quantifies the reduction in uncertainty of one target variable X_{tar} by the knowledge of a source variable X_s. For a pair of random variables (X_s; X_{tar}), MI is given by:

$$I(X_s; X_{\text{tar}}) = \sum p(x_s, x_{\text{tar}})\log_2\left(\frac{p(x_s, x_{\text{tar}})}{p(x_s)p(x_{\text{tar}})}\right) = H(X_{\text{tar}}) - H(X_{\text{tar}} \mid X_s) \tag{3}$$

2.1.3. Multivariate Mutual Information

MMI $I(X_{s1},...X_{sn}; X_{\text{tar}})$ measures the reduction in uncertainty of one target variable X_{tar} by knowledge of multiple source variables. In this study, the simplest case involving two source variables X_{s1} and X_{s2} is examined, of which the function is given as:

$$I(X_{s1}, X_{s2}; X_{\text{tar}}) = \sum p(x_{s1}, x_{s2}, x_{\text{tar}})\log_2\left(\frac{p(x_{s1}, x_{s2}, x_{\text{tar}})}{p(x_{s1}, x_{s2})p(x_{\text{tar}})}\right) \tag{4}$$

2.1.4. Temporal Information

Temporal information is a technique to measure the evolution of dependencies among variables by applying MI or MMI with moving time windows (Goodwell and Kumar 2017a, 2017b). Instead of obtaining a single value MI or MMI for complete time series, temporal information technique cuts the time series into several time windows and obtains MI or MMI for each time window. In a climate application like this, such a measurement is tailored to detect the dependency among environmental factors considering that the dependency could vary due to factors such as seasonality.

2.2. Existing Information Partitioning Approaches

Two partitioning approaches have been proposed in past studies. One is partitioning of information into a linear part and a nonlinear part (Smith, 2015), which is applicable to either MI or MMI. The other is specific to MMI: partitioning of information into components representing the interactive roles of multiple source variables, namely: unique, redundant, and synergistic components (Goodwell and Kumar 2017a, 2017b; Williams & Beer, 2010). Here we describe both approaches and show how they may be combined to decompose information in more detail.

2.2.1. Nonlinearity and Linearity

Total information can be partitioned into linear information and nonlinear information. The method of calculating the nonlinearity in MI has been proposed by Smith (2015). For a given set of X_s and X_{tar}, the procedure is to: (1) fit a linear regression model in terms of predicting X_{tar} given X_s; (2) obtain $\hat{X}_{\text{tar}}$ as the fitted values of X_{tar} and define the nonlinear residual $X'_{\text{tar}} = X_{\text{tar}} - \hat{X}_{\text{tar}}$; (3) normalize X'_{tar} by the quantile normalization based on the value of X_{tar} (quantile normalization makes the distribution of X'_{tar} and X_{tar} identical in statistical properties; see Smith (2015) for a detailed discussion); (4) estimate the MI for both: $I(X_s; X_{\text{tar}})$ and $I\left(X_s; X'_{\text{tar}}\right)$. The quantity $I\left(X_s; X'_{\text{tar}}\right)$ is the nonlinear dependency between X_s and X_{tar}. The linear dependency in terms of MI is the difference (total minus nonlinear): $I\left(X_s; X_{\text{tar}}\right) - I\left(X_s; X'_{\text{tar}}\right)$.

2.2.2. Unique, Redundant, and Synergistic Components

For a system composed of two sources and one target, the total MMI can be decomposed into synergistic, redundant, and two unique components. A unique (U) component is the information shared only between an individual source and the target. Redundancy (R) is repeated information that both sources share with the target. The synergistic (S) component is the extra information arising from the cooperative interaction among the sources. The expression of partitioning is given as the follows:

$$\begin{aligned} I\left(X_{s1}, X_{s2}; X_{\text{tar}}\right) = {} & U_1\left(X_{\text{tar}}; X_{s1}\right) + U_2\left(X_{\text{tar}}; X_{s2}\right) \\ & + R\left(X_{\text{tar}}; X_{s1}, X_{s2}\right) + S\left(X_{\text{tar}}; X_{s1}, X_{s2}\right) \end{aligned} \tag{5}$$

where U_1, U_2, R, and S are nonnegative quantities. Note that each source has its own individual unique contribution, whereas the redundant and synergistic contributions involve both sources.

Using this conceptualization, MI between each source individually and the target can be decomposed as the sum of unique and redundant components:

$$I\left(X_{s1}; X_{\text{tar}}\right) = U_1\left(X_{\text{tar}}; X_{s1}\right) + R\left(X_{\text{tar}}; X_{s1}, X_{s2}\right) \tag{6}$$

$$I\left(X_{s2}; X_{\text{tar}}\right) = U_2\left(X_{\text{tar}}; X_{s2}\right) + R\left(X_{\text{tar}}; X_{s1}, X_{s2}\right) \tag{7}$$

The above equation set (Equations 5–7) requires one to seek an additional equation for any of U_1, U_2, R, and S in order to obtain all components (note that any I can be calculated directly from the probability density function of data). Several ways have been proposed to achieve a well-determined system by estimating U or R, while there is no universal agreement on the best approach. In this study, we use the approach proposed by Goodwell and Kumar (2017a, 2017b), which assumes that the strength of the dependency of the two sources determines the amount of redundant information. To achieve this, a measure called Rescaled Redundancy (R_s) is introduced as the follows:

$$R_s = R_{\min} + I_s\left(R_{\text{MMI}} - R_{\min}\right) \tag{8}$$

The solution of R_s is obtained by computing the normalized source dependency I_s, the lower bounds of redundancy $R_{\min}$, the upper bounds of redundancy R_{MMI}, and the interaction information II ; these are given as follows:

$$I_s = \frac{I\left(X_{s1};X_{s2}\right)}{\min\left[H\left(X_{s1}\right),H\left(X_{s2}\right)\right]} = \frac{I\left(X_{s1};X_{s2}\right)}{\min\left[I\left(X_{s1};X_{s1}\right),I\left(X_{s2};X_{s2}\right)\right]} \tag{9}$$

$$R_{\mathrm{MMI}} = \min\left[I\left(X_{s1};X_{\mathrm{tar}}\right),I\left(X_{s2};X_{\mathrm{tar}}\right)\right] \tag{10}$$

$$R_{\min} = \max\left(0,-II\right) \tag{11}$$

$$II = I\left(X_{s1};X_{s2};X_{\mathrm{tar}}\right) \tag{12}$$

The interaction information II can be either positive or negative and II is shown by Williams and Beer (2010) to be equal to $S - R$. With stronger dependency between the two sources, a larger normalized source dependency I_s results in a larger redundant component R.

2.3. Integrated Information Measurement

Our attempt to quantify the dependency among multiple interacting variables and to disentangle the information as different explainable components is achieved by combining the two approaches with an additional step.

We extend the approach of calculating nonlinearity from the bivariate (MI) to the trivariate (MMI) case. For a given set: X_{S1}, X_{S2}, and X_{tar}, a linear regression model is fitted in terms of predicting X_{tar} given both X_{S1} and X_{S2}. The rest of the procedure to obtain the nonlinear multivariate MI $I\left(X_{s1},X_{s2};X'_{\mathrm{tar}}\right)$ follows as described above. Subsequently, the two partitioning frameworks can be fitted together perfectly. The full decomposition of MMI by the integrated approach results in eight components relating two source variables to one target variable: four linear components $\underline{U}_1$, $\underline{U}_2$, $\underline{R}$, and $\underline{S}$; and four nonlinear components U'_1, U'_2, R', and S'. More precisely, total components U_1, U_2, R, and S are calculated by total MMI decomposition and nonlinear components U'_1, U'_2, R', and S' are calculated by nonlinear MMI decomposition. Then, linear components are calculated by subtracting the nonlinear components from their corresponding total components, for example, $\underline{U}_1 = U_1 - U'_1$.

Although partitioning methods proposed by Smith (2015) and Goodwell and Kumar (2017a, 2017b) both decompose the information into nonnegative components, counter-intuitively, the combination of the two methods does not automatically give four nonnegative linear components and four nonnegative nonlinear components. In practice, nonlinear synergistic information S' computed from the residuals could be larger than its corresponding total synergistic component $\underline{S}$, obtained from the original data. In such cases, we set the nonlinear component to be equal to the total value and reset the linear component to zero. The gap between the integrated approaches and existing information partitioning approaches worth further study to improve ability of information theory to disentangle the complex nature phenomena.

An example of the application on multiple time series is shown in Section 2.5. Validation of this approach is discussed in Supporting Information S1.

2.4. Data

Daily mean fields at a resolution of 1° × 1° are calculated based on UTC dates and times from the Modern-Era Retrospective analysis for Research and Applications, Version 2 (MERRA-2) hourly output (GMAO, 2015) at a resolution of 0.5° × 0.625° spanning 1986–2015. MERRA-2 is a long-term global

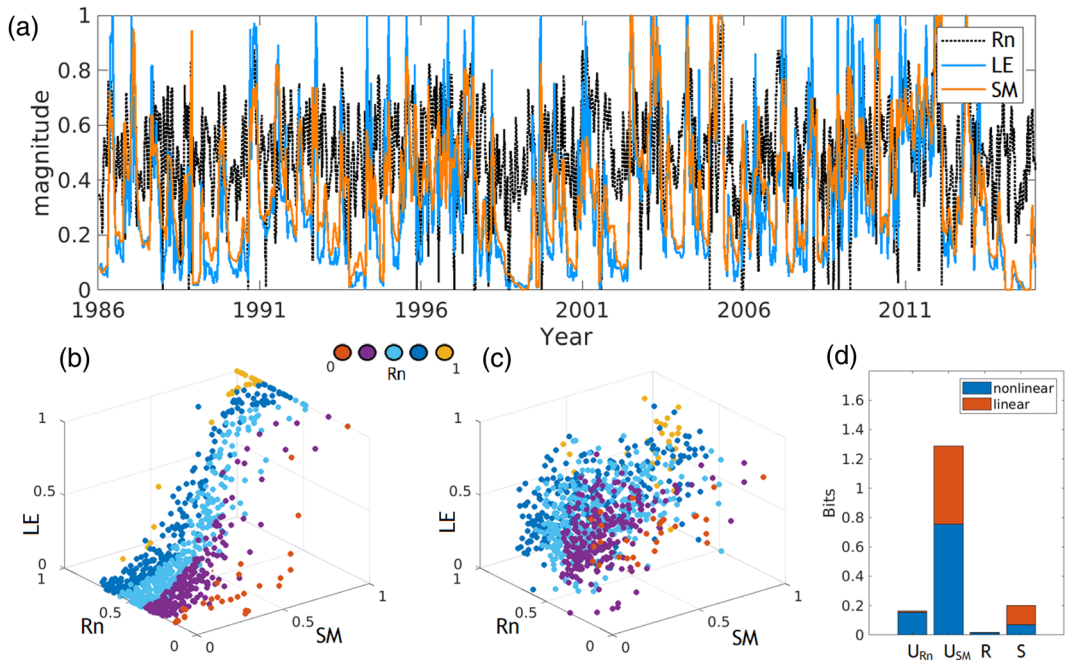

Figure 1. An example of the process of estimating multivariate mutual information (MMI) for a grid cell (10°E, 15°N) during June. The source variables are net radiation (Rn) and soil moisture (SM); latent heat flux (LE) is the target variable. (a) The 30-year June timeseries produced by catenating the 30 years daily data of each June. (b) A 3-d scatter plot of the Rn, SM, and LE used to calculate the total MMI; points are colored by Rn to aid perception. (c) A 3-d scatter plot of the Rn, SM, and the nonlinear residual of LE determined by subtracting the linear model, used to estimate the nonlinearity in total MMI. (d) The decomposition of total MMI into nonlinear and linear parts after further decomposition into unique, redundant, and synergistic components.

reanalysis which provides the depiction of the modern climate. The reanalysis process combines a variety of observations and a forecast model using data assimilation to produce gridded data sets without gaps; the product includes a wide range of variables, part of which are observed irregularly and/or indirectly. Due to its spatial and temporal completeness and consistency, reanalyzes are a powerful scientific tool for climate research. The variables used in this study are not directly measured or assimilated into MERRA-2. However, soil moisture is highly constrained by observed precipitation (Reichle, Draper, et al., 2017; Reichle, Liu, et al., 2017). Net radiation is estimated with information from observed cloud properties and thermodynamic profiles. Surface heat fluxes are calculated with observationally constrained near-surface thermodynamic profiles and winds. The fact that some observed information is used as the "input" of multiple analyzed variables within the MERRA-2 analysis model might lead to some inherent dependencies among the analyzed variables. We explore the information shared by land surface energy and wetness conditions with both latent heat flux and sensible heat flux. Intuitively, total land energy change (net radiation) and soil moisture are used as the sources and the surface heat fluxes are targets.

2.5. Workflow

The climatological seasonal cycle of the coupling strength quantified by MMI is calculated independently at each ice-free land grid cell of MERRA-2. Variability with frequencies lower than 1/365 days is removed by a high-pass filter. Then, for each variable, time series for each calendar month of the 30-year period are constructed. For instance, a catenated time series for June is produced by connecting each June 30th of a year with June 1st of the next year. Then, we use Tukey's fences to deal with the outliers. The values larger than the upper boundary $Q_{max} = Q_3 + 1.5(Q_3 - Q_1)$ are set as Q_{max}; values smaller than the lower boundary $Q_{min} = Q_1 - 1.5(Q_3 - Q_1)$ are set as Q_{min}, where Q_1 and Q_3 are the first and third quartile respectively. Then, each timeseries is normalized into the range [0, 1]; an example is shown in Figure 1a. A 3-D probability

density function (3-D pdf) of the postprocessed data has been calculated with this time series and Equation 4 is applied to calculate the total MMI (Figure 1b). A linear model $\hat{X}_{\text{tar}} = b + \sum_i a_i X_{s_i}$ is then fitted to the time series to calculate the residual of the target X'_{tar} and quantile normalization is applied on X'_{tar} based on the quantile of the target X_{tar} (Figure 1c) to make the total entropy of X'_{tar} and X_{tar} equivalent, ensuring the comparability between total MMI and nonlinear MMI. Again, Equation 4 is applied on this new 3-D pdf to obtain the nonlinear MMI. The decomposition of total MMI is calculated (Figure 1d) by applying Equation 5 to Equation 12. Same calculation is done also for the decomposition of nonlinear MMI. Finally, each linear component is defined as the remainder of its corresponding total component minus its corresponding linear component. Generally, an extreme case with only linear information occurs when (1) variability in X_{tar} is perfectly fitted by a linear combination of X_{si} or (2) X'_{tar} cannot be explained at all by using MMI to quantify the dependency. Situations with identical evolutions of both sources lead to information occupied only by redundancy. Other than that, information provided by either source is an aggregate of unique and redundant components (Equation 6); information provided by the integration of the sources that surpasses the sum of the individual information provided either source is synergy.

A total of eight components are obtained. In this example, variability of latent heat flux is mainly controlled linearly by variability of soil moisture (Figure 1d); this can also be intuited from the 3-D pdf (Figure 1b). Rareness of linear redundancy implies that the part of the uncertainty in latent heat flux linearly explained by soil moisture alone is largely different from the part explained by net radiation alone, that is, soil moisture and net radiation do not covary with each other. The detection of some linear synergy reflects the additional uncertainty in latent heat flux that can be linearly explained by the concurrent action of net radiation and soil moisture, rather than by the sum of individual variations in net radiation and soil moisture. Similar concepts can be used to interpret the nonlinear components. For example, the variability of latent heat flux that cannot explained by a linear model can be partly explained by net radiation and soil moisture (the moderately large nonlinear unique components) and their integration (the nonlinear synergistic component that is as large as the linear synergistic component).

It should be kept in mind when interpreting the results that we only analyze the dependency among co-varying timeseries at daily time scales without lag, and thus the results statistically do not prove casual relationships. Nevertheless, based on known physical processes associated with variations in soil moisture and surface fluxes, it is appropriate to name net radiation and soil moisture as the sources and heat fluxes as the targets. In the Earth system, net radiation is ultimately the main energy source available for the land surface to change the phase of water (the energy released as latent heat flux) or to heat/cool the land surface (the energy released as sensible heat flux) (Trenberth et al., 2009). A large evapotranspiration rate can reduce the amount of soil moisture, contributing to the interdependency between latent heat flux and soil moisture. However, at daily time scales, except for regions with abundant soil moisture where latent heat flux is not moisture limited, soil moisture plays the key role in controlling the partitioning of the surface heat fluxes (Seneviratne et al., 2010).

Note that because the MMI is calculated separately for each calendar month, subtracting the monthly climatology from the data does not affect the distribution (i.e., additional filtering out of the seasonal cycle is not necessary). Besides, the month-to-month analyses enables us to avoid the arbitrariness in the selection of an approach to remove seasonality. Seasonality leads to inherent dependencies among many meteorological variables that are not the focus of this study. For example, 3-month time-series could be largely correlated because the same location on the Earth receives a regularly varying solar radiation that dominates the behavior of many meteorological variables at a seasonal time scale. Although seasonality can be removed by applying a high-pass filter, estimating a climatological seasonal cycle and subtracting, or removing specific harmonics in the time series, we argue that any of those could introduce artificial dependencies in the results. An information theory-related algorithm completely utilizes data structures without loss of information or dimensionality reduction (Perdigão et al., 2020). The mentioned methods can alter the data distribution in a different way and thus leads to different MI.

The unit of MMI is bits, representing the amount of information transmitted from the sources to the target. To make the result more interpretable, obtained MMI values are normalized by the Shannon entropy of

the target $H(X_{tar})$. Therefore, the amount of normalized MMI (nMMI) can be interpreted as the fraction of uncertainty of the target that is explained by the sources.

2.6. Significance Testing

We test the statistical significance progressively for total MMI, nonlinear MMI, and linear MMI. If a total MMI quantity is found to be significant, full decomposition of both the linear and nonlinear parts of the unique, redundant and synergistic components is computed. The term "preprocessed timeseries" means the timeseries that have undergone high-pass filtering, Tukey's fences, and normalization mentioned in Section 2.4.

For the total MMI case, a shuffled surrogates method, with the null hypothesis that no total dependency exists, is applied on each grid cell and each calendar month. Once we calculate the MMI from the preprocessed timeseries (observed MMI), we resample the timeseries by randomly permuting the preprocessed timeseries of each of the two sources and calculate MMI again. By repeating the process 100 times, a probability distribution of MMI as well as its mean μ and standard deviation σ are obtained. We retain the observed MMI that is larger than $\mu + 3\sigma$, the level of 99% confidence. A fully nonparametric significance threshold can be directly obtained by repeating the process ~1,000 times, but it is much more computationally expensive while yielding very similar results when tested on individual grid cells.

An identical procedure and null hypothesis are used to calculate the significance of nonlinear MMI. The observed nonlinear MMI and the MMI computed from the shuffled surrogates method both use the target that has had its linear fit subtracted. An observed nonlinear MMI larger than $\mu + 3\sigma$ means the dependence is significant at the 99% confidence level and it can be recognized that only nonlinearity occupies such significant total dependency.

For the linear component, we first find the 99% significance value of the multiple correlation coefficient ρ_c for the given three preprocessed timeseries. Then, the criterion for linear MMI, MMI_c, can be calculated by using the following equivalence between correlation and MI under the assumption that variables have a Gaussian distribution:

$$\mathrm{MMI}_c = -\frac{1}{2}\log\left(1-\rho_c^2\right) \tag{13}$$

3. Results

The seasonal cycle of total nMMI (each season is reported as the mean of three analyzed months), quantifying the dependency of latent heat (LE) on land surface net radiation (Rn) and soil moisture (SM), is shown in Figure 2. Regions where the surface temperature is below 0°C for more than half of the days during the analyzed season are masked out and regions where all months do not pass the significance test are shaded gray. The total dependency over most of the world is significant. This is not a surprising result since it has long been recognized that latent heat flux is controlled mainly by available moisture and energy (Seneviratne et al., 2010; Trenberth et al., 2009). The strongest dependencies are found mainly over tropical areas during all seasons. Semiarid regions and the Asian monsoon area show large values during wet seasons (Figure 2c). Dependencies using sensible heat flux (H) as the target are mostly smaller over the globe as shown in Figure 3. Tropical rainforest in this case has much lower total nMMI while the large values over semiarid regions during wet seasons remain.

The nonlinear and linear components contributing to the total nMMI for LE and H as the target are shown in Figures 4 and 5 respectively. To make the sum of nonlinear and linear components equal to the total normalized MMI, we only screen out the regions when none of analyzed months are statistically significant. Except for high latitudes, most areas have moderate to large nonlinearity. Some particularly large values are found over arid regions, for example over the Sahara and the Arabian Peninsula in both seasons and over Western Australia in December–January–February (DJF). In those desert regions, soil moisture content is usually below the critical (wilting) point (Agam (Ninari) et al., 2004) so that the wetness conditions of the

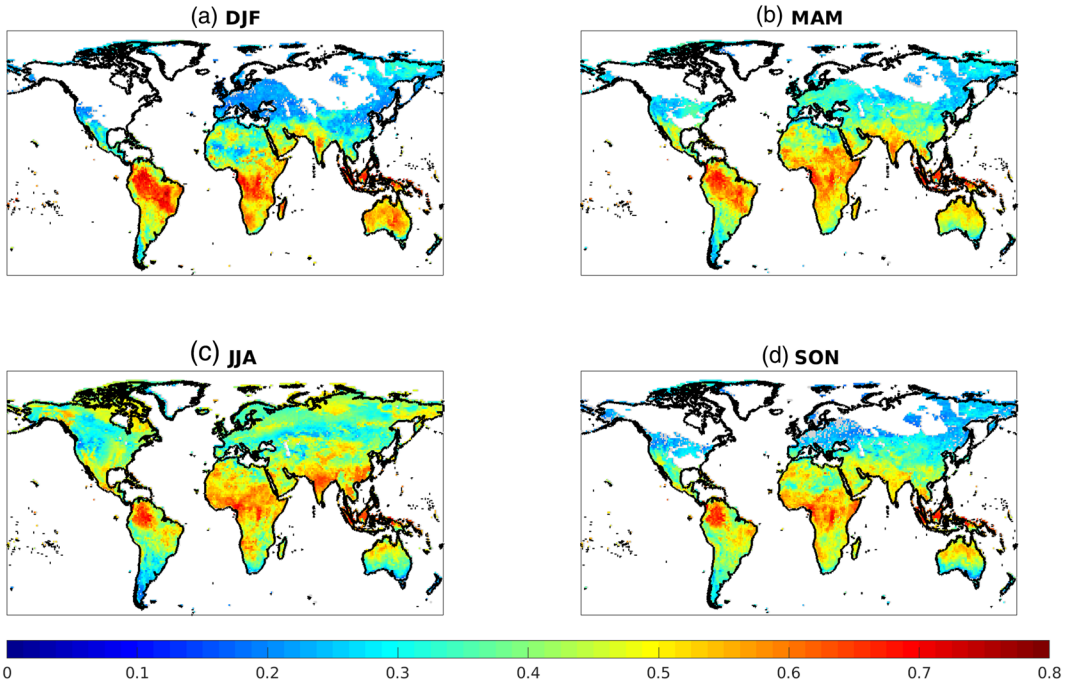

Figure 2. Seasonal total normalized multivariate mutual information (nMMI, unitless) with latent heat flux (LE) as the target, net radiation (Rn) and soil moisture (SM) as the sources. Gray shading means that not all three analyzed months pass the significance test. Blank land grid cells are regions where the surface temperature is below 0°C for more than half of the analyzed period.

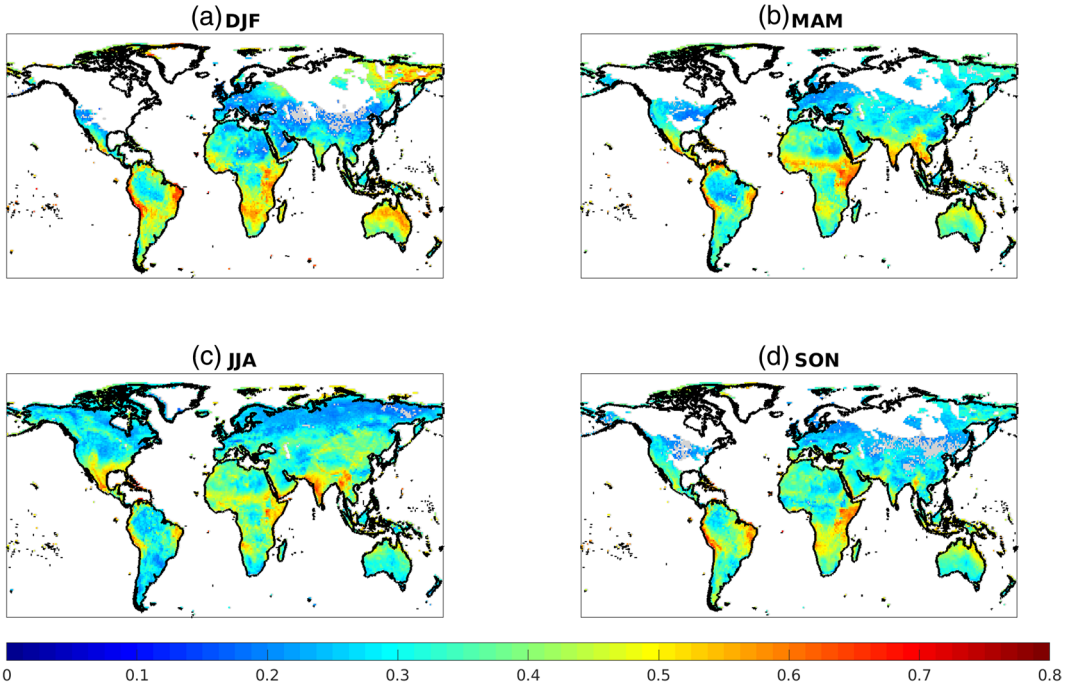

Figure 3. As in Figure 2 but using sensible heat flux (H) as the target.

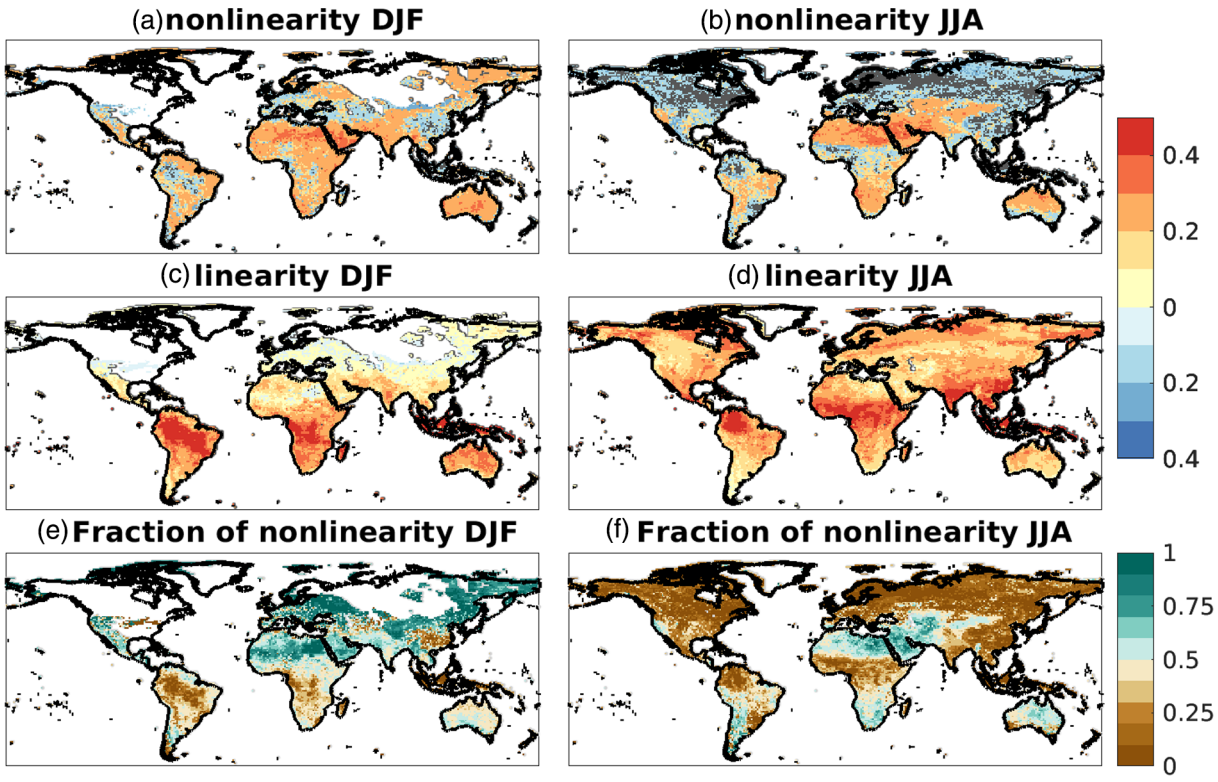

Figure 4. During DJF: (a) Nonlinear nMMI, (c) linear nMMI, and (e) the fraction of nonlinearity to total nMMI with LE as the target and Rn and SM as the sources. (b), (d), and (f) are same as (a), (c), and (e) respectively but for JJA. In panels (a) to (d), different color palettes indicate whether all months of the season passed the significant test: areas where no analyzed months pass the significance test are shaded dark gray; areas where all analyzed months pass the significance test are shaded in warm colors; areas where one or two analyzed months pass the significance test are shaded in cool colors. Blank land grid cells are regions where the surface temperature is below 0°C for more than half of the analyzed period. All quantities are unitless. DJF, December–January–February.

land do not affect the release of latent heat flux. However, once soil moisture content rises high enough, latent heat flux becomes sensitive to the change in soil moisture. Such a transition is infrequent, but it can induce a dramatic change in the relationship between soil moisture and latent heat flux, resulting in large nonlinearity over those areas. Linearity is strong over wet and semi-arid regions (Figures 4c and 4d) and has a strong north-south gradient during DJF (Figure 4c). The fraction of nonlinearity shows there is generally strong linear dependency over the summer hemisphere, while nonlinear contributions are found to be more important over the winter hemisphere. However, arid and some semi-arid and subtropical regions show a substantial fraction of nonlinear contribution throughout the year.

Results using H as the target (Figure 5) show a slightly different pattern. The nonlinearity is more homogeneous and subdued than in the LE case. Semiarid regions including Mexico, the Sahel, and the Indus Valley during JJA, northern and eastern Australia, the South American lowlands, and southern Africa during DJF are clearly the standouts of linear dependency. The presence of strong nonlinearity in nMMI for H is largely absent over deserts, and is large only over parts of the tropics, namely in local dry seasons (Figures 5e and 5f), and in the Northern Hemisphere mid-latitudes during DJF (Figure 5e).

We next focus on JJA to examine the composition of both the linear and nonlinear dependencies. The decomposition of total, linear, and nonlinear nMMI into unique, redundant and synergistic components with LE as the target is displayed in Figure 6. The linear unique contributions from SM alone (Figure 6b), shows a pattern similar to the canonical land-atmosphere "hot spots," which are detected mostly over semi-arid regions such as the Sahel (Dirmeyer, 2011; Koster et al., 2004), while the dry regions also depict large dependencies between LE and SM in our analysis. Ignoring the magnitude of change in the variable, which is a considered factor for quantifying coupling strength in past studies (e.g., the standard deviation term in Guo et al., 2006), might lead to this difference in the pattern. By comparing Figures 6f and 6j, we see many

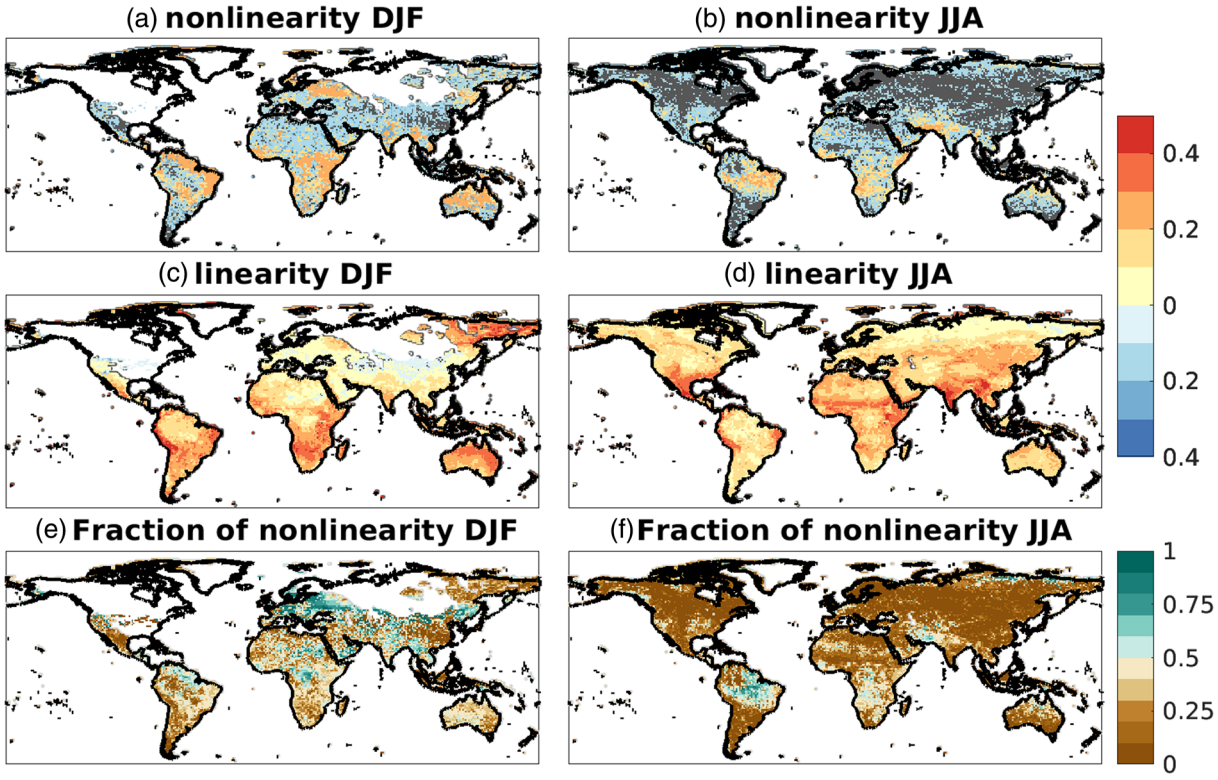

Figure 5. As in Figure 4 but using H as the target.

of the areas with strong linear unique contributions from SM also have little nonlinear SM-LE dependency (Figure 6f). This can be attributed to the threshold behavior, characterized by SM values distributed around the wilting point or critical point, and/or the higher order direct relationships within the transitional zone of SM-LE relationships. We note that nonlinearity is prominent in much of the semiarid area with both of the above-mentioned features (not shown) and their quantification needs further investigation. Such non-negligible contribution of nonlinearity by SM suggests that quantifying the coupling under a linear framework, as in past studies (Dirmeyer, 2011; Koster et al., 2004), may underestimate the strength and somewhat misrepresent the character of coupling over the "hot spots." The case using H as the target (Figures 7f and 7j) shows somewhat similar patterns as that using LE, while the strength of dependencies is weaker overall, and no strong dependency is found over dry regions.

Comparison between the linear unique contributions from Rn and SM (Figures 6i and 6j) suggests the two patterns are largely out of phase, accompanied by very weak linear redundancy (Figure 6k). They reveal that the two dominant regimes are controlled solely by either energy or moisture. Such a bimodal pattern is evident even when the nonlinearity is included (Figures 6a–6d). This validates previous studies that divide the globe into energy-limited regions and soil moisture-limited regions when only considering the linear dependencies (e.g., Teuling et al., 2009). The nonlinear contribution from Rn alone (Figure 6e) is nonzero but much weaker than that from SM (Figure 6f); no bimodal pattern is found as is seen in the linear part (Figures 6i and 6j).

Whereas there is almost no redundancy between SM and Rn (Figure 6c), some degree of synergistic information (Figure 6d) is found over much of the world. The nonlinear synergistic information (Figure 6h) is much larger than the linear part (Figure 6i). This suggests the linear SM-LE relationship is not obviously modulated by Rn; a reason that could lead to a neglect of any multidimensional SM-Rn-LE relationship by statistical frameworks with linear dependencies. Large nonlinear and linear synergistic components are found over many of the semiarid regions, for example, the Sahel, India, and northern China. In these regions, the soil moisture content typically lies in the transition zone wherein LE is sensitive to fluctuations in SM. Together with the large synergistic information, this result suggests that both the linear and nonlinear

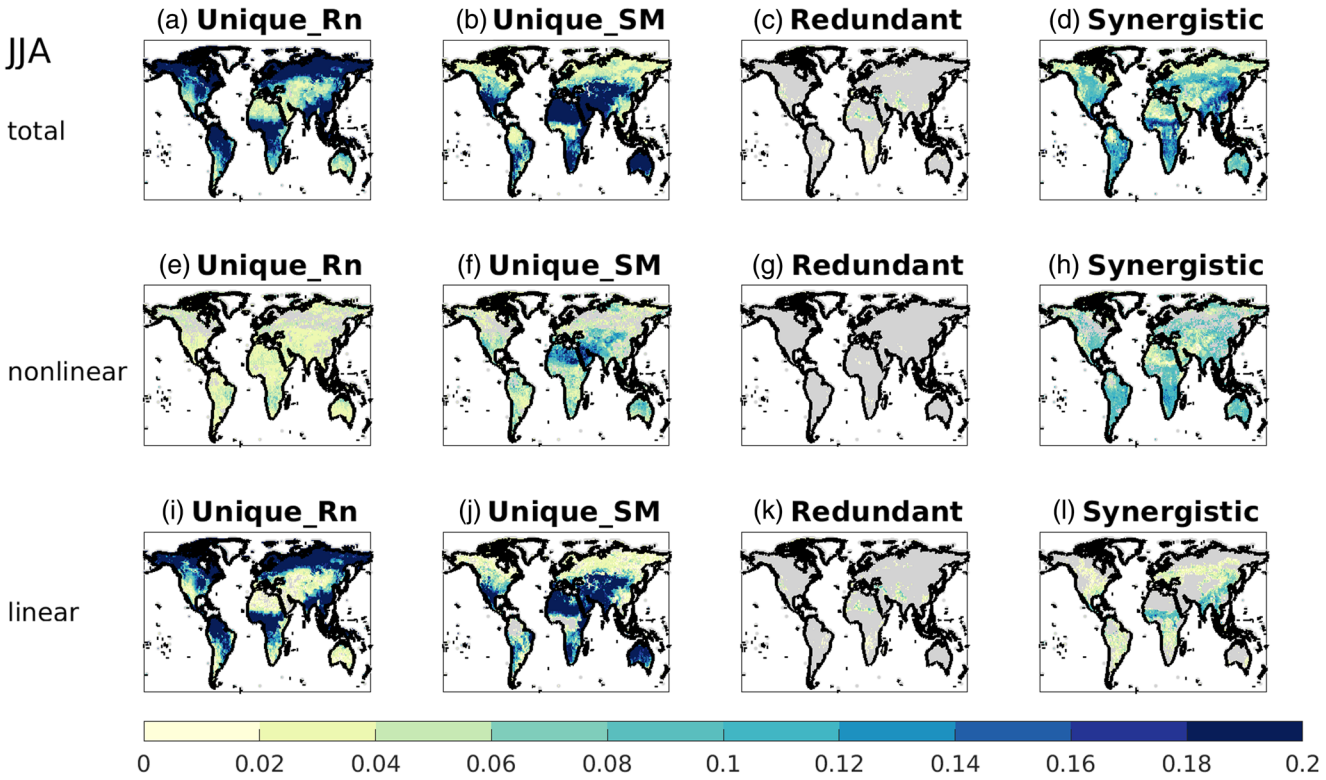

Figure 6. The partitioning of total nMMI with LE as the target, Rn and SM as the sources for boreal summer (JJA); decomposition is calculated over regions of where nMMI is significant for at least one of the three analyzed months. (a) unique information contributed from Rn, (b) unique information contributed from SM, (c) redundant and (d) synergistic components. (e–h) are same as (a–d) respectively but for the partitioning of the nonlinear nMMI. (i–l) are same as (a–d) respectively but for the partitioning of the linear nMMI. All the quantities are unitless. Regions where decomposition is not calculated or where the value is less than 0.01 are shaded gray. nMMI, normalized multivariate mutual information.

relationships between soil moisture and heat fluxes can be modulated by Rn. Our result corroborates the findings in a recent observational station-based analysis that the relationship between soil moisture and heat fluxes is multidimensional (Haghighi et al., 2018). However, here we demonstrate that such a multidimensional concept applies globally.

The bimodal pattern seen between Rn and SM sources for LE is less distinct for H (Figures 7a, 7b, 7i, and 7j), in which the linear contribution of Rn dominates the total nMMI and is opposite to that in LE case in many moisture-limited regions, for example western North America, the Sahara and the Arabian Peninsula. Intuitively, in those regions, the available energy directly determines the amount of H most of the time since there is no water to be evaporated; the available moisture determines the amount of LE only when soil moisture content is above a critical value. The nonlinear contribution from SM alone (Figure 7f) is much weaker compared to the case of LE (Figure 6f) and no particularly large value is found. This suggests that when soil moisture passes the critical point, the induced change in the SM-H relation is not as obvious as that in the SM-LE relation. The nonlinear synergistic contribution and linear contribution of SM are comparable, again suggesting the nonlinear and multidimensional dependencies among Rn, SM, and H.

The appearance of bimodal patterns for LE and to a lesser extent H gives rise to the question of which source dominates the partitioning of surface heat fluxes. To address this, a decomposition with the target of evaporative fraction (EF), calculated as LE divided by the sum of LE and H, is shown in Figure 8. The decomposition of MMI reveals that SM plays a critical role in heat flux partitioning; most of the contribution is linear, while the nonlinearity is slightly larger than that in both the LE and H cases. The total synergistic information is larger than the unique contribution from Rn, suggesting that the SM-EF relationship can be affected by Rn, although the Rn-EF relationship is weaker. This highlights the importance of exploring coupling in a multivariate analysis. We note that the nonlinear contribution to the total nMMI in the EF case (not shown) is greater than in the separate LE and H cases, and could arise from the mathematical representation of

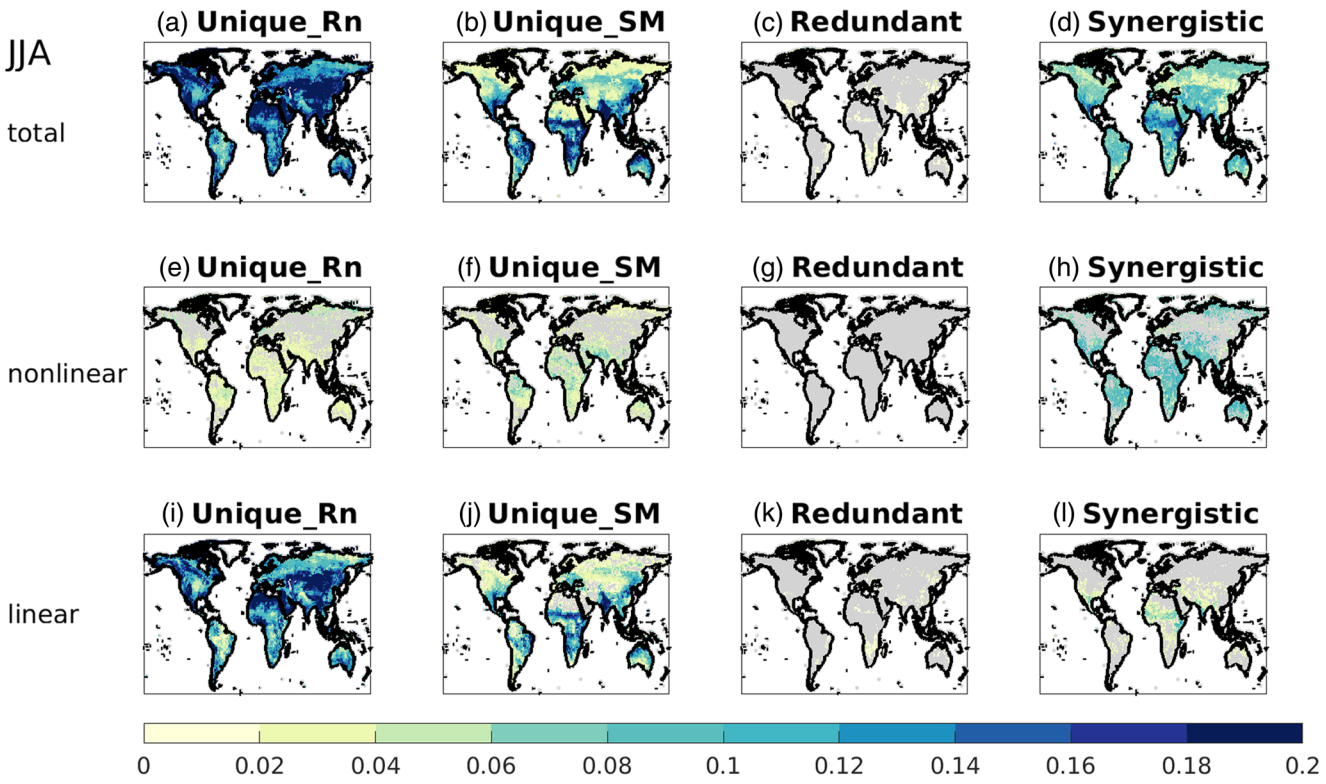

Figure 7. Same as Figure 5 but using H as the target.

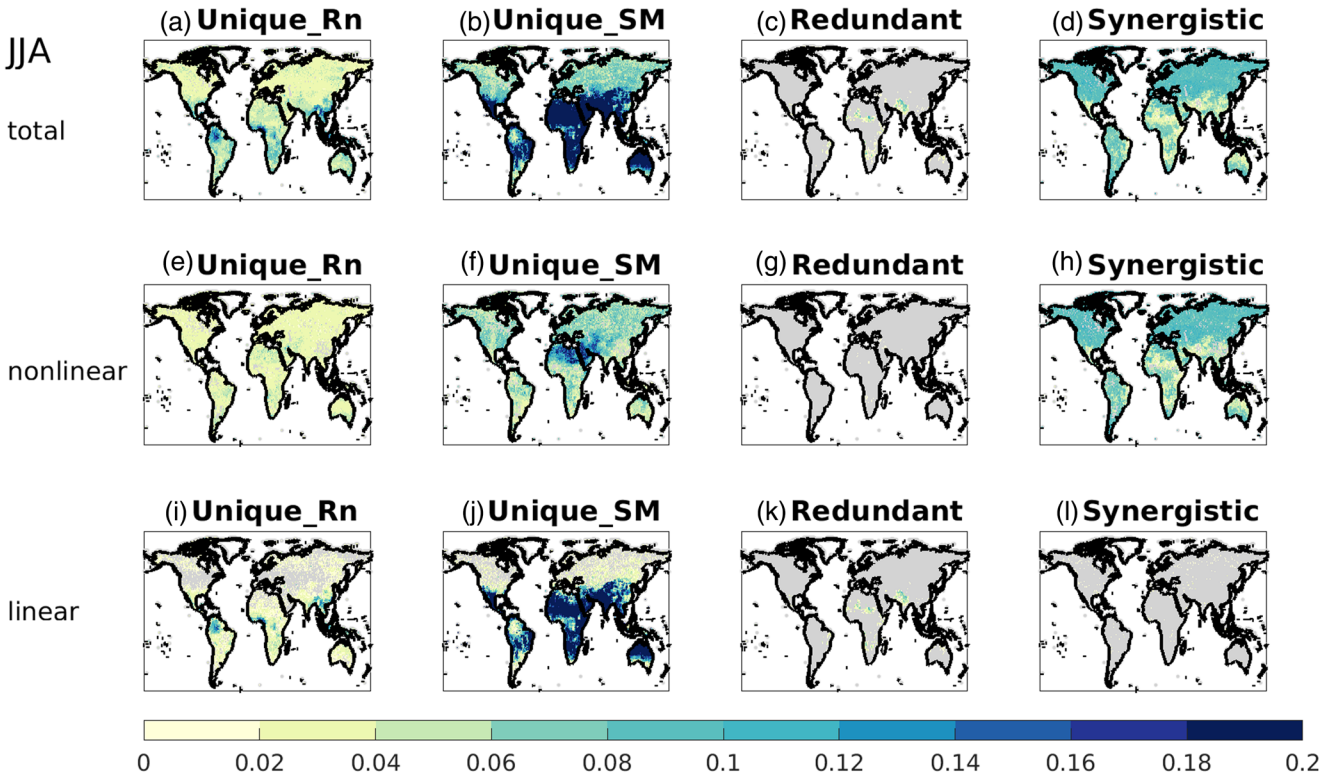

Figure 8. Same as Figure 5 but using evaporative fraction (EF) as the target.

the EF ratio. Individually, linear SM-LE and SM-H relationships can still result in a nonlinear SM-EF relationship as the ratio form of EF makes it inherently sensitive to fluctuations in the denominator when it is small. In high latitudes or places with cloud cover, such situations can occur because net radiation is small or even negative, leading to values of EF well outside the nominal range 0–1. After EF is calculated from the original data of MERRA-2, we rely entirely on Tukey's fences so that the preprocessed timeseries of EF is constrained to the range [0, 1].

Overall, we see that the unique components for LE clearly partition such that SM is a dominant source in water-limited regions, and Rn dominates in energy-limited regions, whereas for H it is Rn that dominates almost everywhere. Patterns for synthesis are more similar between LE and H, but generally stronger for LE. Both targets appear to have nonlinear synergistic components that are widespread and account for around 10% of the total contribution. The linear synergistic component is largely confined to monsoonal regions for both LE and H. It is interesting that the unique linear SM contribution to H as well as the linear synergistic contributions to both LE and H are strongest in transitional regions between arid and humid climates, which are the canonical hot spots of land-atmosphere coupling (Koster et al., 2004), yet the synergistic component for EF lacks spatial structure and contains virtually no linear component. None of the surface heat flux terms appear to have a notable redundant component among sources.

4. Conclusions

Addressing both nonlinear and multidimensional aspects, a technique based on information theory has been applied to reanalysis data to revisit the global estimation of land-atmosphere coupling. A quantification of multidependency using MMI, has been decomposed as different interpretable components by a newly proposed integrated partitioning method. Three combinations of three variables (two sources, one target) have been explored, namely Rn-SM-LE, Rn-SM-H, and Rn-SM-EF, of which Rn and SM are the sources and their contributions to the change in targets LE, H, and EF have been quantified respectively.

Our analysis of total multidimensional dependency shows variability in both spatial and temporal aspects. The linear components resemble the canonical land-atmospheric coupling distributions found in previous studies as well as demarking the regions known to be governed by water-limited and energy-limited regimes. The nonlinear components superposed on the linear results do not alter these familiar patterns and thus strengthen the authoritativeness of past findings, while contributing new insights. Nonlinear contributions to LE variability are predominant in arid regions and across midlatitude and subtropical areas in the winter hemisphere. Most of the nonlinear contribution is from SM, although there are nonnegligible contributions from Rn. The existence of water- and energy-dominated regimes seen for LE are also evident for H. Whereas Rn is a major contributor to variability in LE and H individually over much of the world, the partitioning of surface heat fluxes is confirmed to be strongly determined by SM by using EF as the target variable. The property of multidimensional dependency among land-atmospheric coupling factors is revealed to exist over the whole globe by the substantial magnitude of the synergistic term, which is greater than the redundancy term in all cases.

We have only applied this analysis to MERRA-2, which is not a perfect representation of reality since values of net radiation and surface heat fluxes are calculated within an Earth system model from other assimilated state variables rather than being measured directly. This leads to an inherent interdependency among the variables analyzed in this study. Further application of our analysis on data from other reanalyzes, climate models, and satellite data are needed to increase confidence in the global patterns shown here.

We also note that the nonlinearity and synergistic relationships suggested to exist across the globe need further investigation. For instance, it remains to be disentangled how the contribution to such nonlinear relationships arise from factors including (1) potential critical points that determine the changes in sensitivity of surface heat fluxes to soil moisture, (2) higher order direct relationships between variables, and (3) natural groupings inherent in the data. Synergism, treated as the ability of a third factor to alter the bivariate relationship between two factors, is worth quantifying to improve our understanding of the interactions in nature and advance realism and predictability in models. For example, the finding of multidimensional relationships among Rn-SM-H implies that considering Rn as a predictor could improve forecasts of extreme events like heatwaves, since soil moisture-sensible heat flux-surface temperature feedbacks are already known to play a crucial role in predicting near surface temperature.

Finally, we note that the MMI analysis can be performed with other combinations of source variables such as wind speed and near surface humidity. The comparison among different combinations of source variables may further determine when and where other variables not considered in this study are also important factors for surface heat fluxes. Applying this analysis to outputs from numerical models can help identify shortcomings in the parameterizations of land surface processes and land-atmosphere interactions. In addition, the MMI technique can be extended vertically along the water and energy cycle process chains linking land and atmosphere (Santanello et al., 2018) by using surface heat fluxes as the sources and any property/state of the planetary boundary layer, clouds or precipitation as targets. Overall, MMI is a tool that shows great promise for exploring more complex relationships in coupled land-atmosphere processes than have been possible with simple statistics.

Data Availability Statement

MERRA-2 hourly output used in this study were downloaded from Global Modeling and Assimilation Office (GMAO 2015, https://doi.org/10.5067/RKPHT8KC1Y1T).

Acknowledgments
We are grateful to Dr. Timothy DelSole for helpful discussions during key phases of the development of this methodology. This research has been supported by grants from the National Science Foundation (AGS-1419445) and the National Oceanographic and Atmospheric Administration (NA16OAR4310095).

References

Agam (Ninari), N., Berliner, P. R., Zangvil, A., & Ben-Dor, E. (2004). Soil water evaporation during the dry season in an arid zone. *Journal of Geophysical Research*, *109*, D16103. https://doi.org/10.1029/2004JD004802

Budyko, M. I. (1963). *Evaporation under natural conditions, Isr. Program for Sci. Transl., Off. of Techn. Serv*. Jerusalem: U.S. Dep. of Commerce.

Budyko, M. I. (1974). *Climate and life, xvii* (p. 508). ScienceDirect. ISBN: 9780080954530.

Cover, T. M., & Thomas, J. A. (2006). *Elements of information theory* (2nd ed.). Wiley-Interscience.

Dirmeyer, P. A. (2011). The terrestrial segment of soil moisture-climate coupling. *Geophysical Research Letters*, *38*, L16702. https://doi.org/10.1029/2011GL048268

Eltahir, E. A. B. (1998). A soil moisture–rainfall feedback mechanism: 1. Theory and observations. *Water Resources Research*, *34*, 765–776. https://doi.org/10.1029/97WR03499

Global Modeling and Assimilation Office (GMAO). (2015). *MERRA-2 tavg1_2d_lnd_Nx: 2d,1-hourly, time-averaged, single-level, assimilation, land surface diagnostics V5.12.4*. Greenbelt, MD: Goddard Earth Sciences Data and Information Services Center (GES DISC). https://doi.org/10.5067/RKPHT8KC1Y1T

Goodwell, A. E., & Kumar, P. (2017a). Temporal Information Partitioning Networks (TIPNets): A process network approach to infer ecohydrologic shifts. *Water Resources Research*, *53*, 5899–5919. https://doi.org/10.1002/2016WR020218

Goodwell, A. E., & Kumar, P. (2017b). Temporal information partitioning: Characterizing synergy, uniqueness, and redundancy in interacting environmental variables. *Water Resources Research*, *53*, 5920–5942. https://doi.org/10.1002/2016WR020216

Guo, Z., Dirmeyer, P. A., Koster, R. D., Sud, Y. C., Bonan, G., Oleson, K. W., et al. (2006). GLACE: The global land-atmosphere coupling experiment. 2. Analysis. *Journal of Hydrometeorology*, *7*, 611–625. https://doi.org/10.1175/JHM511.1

Haghighi, E., Short Gianotti, D. J., Akbar, R., Salvucci, G. D., & Entekhabi, D. (2018). Soil and atmospheric controls on the land surface energy balance: a generalized framework for distinguishing moisture-limited and energy-limited evaporation eegimes. *Water Resources Research*, *54*, 1831–1851. https://doi.org/10.1002/2017WR021729

Herold, N., Kala, J., & Alexander, L. V. (2016). The influence of soil moisture deficits on Australian heatwaves. *Environmental Research Letters*, *11*, 1–8. http://dx.doi.org/10.1088/1748-9326/11/6/064003

Kala, J., Evans, J. P., & Pitman, A. J. (2015). Influence of antecedent soil moisture conditions on the synoptic meteorology of the Black Saturday bushfire event in southeast Australia. *Quarterly Journal of the Royal Meteorological Society*, *141*, 3118–3129. https://doi.org/10.1002/joc.2206

Koster, R. D., Dirmeyer, P. A., Guo, Z., Bonan, G., Chan, E., Cox, P., et al. (2004). Regions of strong coupling between soil moisture and precipitation. *Science*, *305*, 1138–1140. https://doi.org/10.1126/science.1100217

Los, L. O., Weedon, G. P., North, P. R. J., Kaduk, J. D., Taylor, C. M., & Cox, P. M. (2006). An observation-based estimate of the strength of rainfall-vegetation interactions in the Sahel. *Geophysical Research Letters*, *33*, L–16402. https://doi.org/10.1029/2006GL027065

Neuper, M., & Ehret, U. (2019). Quantitative precipitation estimation with weather radar using a data- and information-based approach. *Hydrology and Earth System Sciences*, *23*(9), 3711–3733. https://doi.org/10.5194/hess-23-3711-2019

Perdigão, R. A. P., U Ehret, U., Knuth, K. H., & Wang, J. (2020). Debates: does information theory provide a new paradigm for earth science? Emerging concepts and pathways of information physics. *Water Resources Research*, *56*(2), 1–13. https://doi.org/10.1029/2019WR025270

Qiu, J., Crow, W. T., Dong, J., & Nearing, G. S. (2020). Model representation of the coupling between evapotranspiration and soil water content at different depths. *Hydrology and Earth System Sciences*, *24*, 581–594. https://doi.org/10.5194/hess-24-581-2020

Reichle, R. H., Draper, C. S., Liu, Q., Girotto, M., Mahanama, S. P. P., Koster, R. D., & De Lannoy, G. J. M. (2017). Assessment of MERRA-2 land surface hydrology estimates. *Journal of Climate*, *30*, 2937–2960. https://doi.org/10.1175/JCLI-D-16-0720.1

Reichle, R. H., Liu, Q., Koster, R. D., Draper, C. S., Mahanama, S. P. P., & Partyka, G. S. (2017). Land surface precipitation in MERRA-2. *Journal of Climate*, *30*, 1643–1664. https://doi.org/10.1175/JCLI-D-16-0570.1

Ruddell, B. L., & Kumar, P. (2009). Ecohydrologic process networks: 1. Identification. *Water Resources Research*, *45*, W03419. https://doi.org/10.1029/2008WR007279

Santanello, J. A., Dirmeyer, P. A., Ferguson, C. R., Findell, K. L., Tawfik, A. B., Berg, A., et al. (2018). Land–atmosphere interactions: the LoCo perspective. *Bulletin of the American Meteorological Society*, *99*, 1253–1272. https://doi.org/10.1175/BAM/S-D-17-0001.1

Santanello, J. A., Peters-Lidard, C. D., Kennedy, A., & Kumar, S. V. (2013). Diagnosing the nature of land–atmosphere coupling: A case study of dry/wet extremes in the U.S. Southern Great Plains. *Journal of Hydrometeorology*, *14*, 3–24. https://doi.org/10.1175/JHM-D-12-023.1

Seneviratne, S. I., Corti, T., Davin, E. L., Hirschi, M., Jaeger, E. B., Lehner, I., et al. (2010). Investigating soil moisture-climate interactions in a changing climate: A review. *Earth-Science Reviews*, *99*(3–4), 125–161. https://doi.org/10.1016/j.earscirev.2010.02.004

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*, 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Smith, R. (2015). A mutual information approach to calculating nonlinearity. *The ISI's Journal for the Rapid Dissemination of Statistics Research*, *4*, 291–303. https://doi.org/10.1002/sta4.96

Tao, C., Zhang, Y., Tang, S., Tang, Q., Ma, H.-Y., Xie, S., & Zhang, M. (2019). Regional moisture budget andland-atmosphere coupling over the U.S. Southern Great Plains inferred from the ARM long-term observations. *Journal of Geophysical Research: Atmospheres*, *124*, 10091–10108. https://doi.org/10.1029/2019JD030585

Teuling, A. J., Hirschi, M., Ohmura, A., Wild, M., Reichstein, M., Ciais, P., et al. (2009). A regional perspective on trends in continental evaporation. *Geophysical Research Letters*, *36*, L02404. https://doi.org/10.1029/2008GL036584

Trenberth, K. E., Fasullo, J., & Kiehl, J. (2009). Earth's global energy budget. *Bulletin of the American Meteorological Society*, *90*(3), 311–323. https://doi.org/10.1175/2008BAMS2634.1

Williams, P. L., & Beer, R. D. (2010). *Nonnegative decomposition of multivariate information*. arXiv, arXiv:1004.2515.

Yu, Y., Notaro, M., Wang, F., Mao, J., Shi, X., & Wei, Y. (2017). Observed positive vegetation-rainfall feedbacks in the Sahel dominated by a moisture recycling mechanism. *Nature Communications*, *8*, 1873. https://doi.org/10.1038/s41467-017-02021-1

References From the Supporting Information

Paninski, L. (2003). Estimation of entropy and mutual information. *Neural Computation*, *15*(6), 1191–1253. https://doi.org/10.1162/089976603321780272